

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BOARD OF PATENT APPEALS AND INTERFERENCES**

In re patent application of:
Campbell et al. Atty. Docket No.: ARC920030088US1

Serial No.: 10/723,112 Group Art Unit: 2162

Filed: November 26, 2003 Examiner: Colan, Giovanna B.

For: SYSTEM AND METHOD FOR RETRIEVING DOCUMENTS OR SUB-DOCUMENTS
BASED ON EXAMPLES

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPELLANTS' APPEAL BRIEF

Sir:

Appellant respectfully appeals the final rejection of claims 1-57, in the Office Action dated October 19, 2006. A Notice of Appeal was timely filed on January 19, 2007.

I. REAL PARTY IN INTEREST

The real party in interest is International Business Machines Corporation, Armonk, New York, assignee of 100% interest of the above-referenced patent application.

II. RELATED APPEALS AND INTERFERENCES

There are no other appeals or interferences known to Appellants, Appellants' legal representative or Assignee which would directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

III. STATUS OF CLAIMS

Claims 1-57, all the claims pending in the application and set forth fully in the attached appendix (Section IX), are under appeal. Claims 1-57 were originally filed in the application. A non-final Office Action was issued on May 19, 2006 rejecting claims 1-57. The Appellants filed an Amendment under 37 C.F.R. §1.111 on August 7, 2006 amending claims 1, 20, and 39. A final Office Action was issued on October 19, 2006 rejecting claims 1-57. The Appellants filed an Response under 37 C.F.R. §1.116 on December 18, 2006. An Advisory Action was issued on January 19, 2007 indicating that the Response filed under 37 C.F.R. §1.116 on December 18, 2006 does not place the application in condition for allowance. The Appellants filed a Notice of Appeal timely on January 19, 2007.

Claims 1-57 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Palmer, et al. (U.S. Patent No. 6,990,628 B1), hereinafter referred to as "Palmer" in view of Woo (U.S. Patent No. 7,039,641 B2). Appellants respectfully traverse these rejections based on the following discussion.

IV. STATEMENT OF AMENDMENTS

A final Office Action dated October 19, 2006 stated all the pending claims 1-57 were rejected. The claims shown in the appendix (Section IX) are shown in their amended form as of the August 7, 2006 amendment.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The Appellants' claimed invention is generally described in pages 5 through 26 of the specification and shown in Figures 1 through 6 of the application as originally filed. More particularly, with reference to the claimed subject matter (with specific reference to the page/line numbers of the Appellants' specification and figures, as originally filed, given in parenthesis):

Claim 1: A system (page 23, line 18, Figure 5, element number 200) for extracting information comprising: a query input (page 23, line 18, Figure 5, element number 201); a database (page 23, line 19, Figure 5, element number 203) of documents (page 23, line 19, Figure 5, element number 210); a plurality of classifiers (page 23, line 19, Figure 5, element numbers 202₁,...,202_n) arranged in a hierarchical cascade (page 23, line 20, Figure 5, element number 202) of classifier layers (page 23, line 20, Figure 5, element number 204), wherein each classifier (page 23, lines 20, Figure 5, element numbers 202₁,...,202_n) comprises a set of weighted training data points comprising feature vectors representing any portion of a document (page 23, lines 21-22, Figure 5, element number 210), wherein each said feature vector is arranged only as a vector of counts for all features in a data point (page 11, lines 12-18), and wherein said classifiers are operable to retrieve documents from said database based solely on whether said documents are

relevant to said query input (page 23, lines 22-23; page 24, line 15 to page 25, line 3); and a terminal classifier (page 24, line 1, Figure 5, element number 205) weighing an output from said cascade according to a rate of success of query terms being matched by each layer of said cascade (page 24, lines 1-2, Figure 5).

Claim 2: The system of claim 1, wherein each classifier accepts an input distribution of data points and transforms said input distribution to an output distribution of said data points (page 10, lines 4-6).

Claim 3: The system of claim 2, wherein each classifier is trained by weighing training data points at each classifier layer in said cascade by an output distribution generated by each previous classifier layer, and wherein weights of said training data points of said first classifier layer are uniform (page 10, lines 6-9).

Claim 4: The system of claim 3, wherein each classifier is trained according to said query input (page 10, line 9).

Claim 5: The system of claim 2, wherein said query input is based on a minimum number of example documents (page 10, lines 9-10).

Claim 6: The system of claim 1, wherein said document comprises data points comprising feature vectors representing any portion of said document (page 9, lines 21-22).

Claim 7: The system of claim 1, wherein said documents comprise a file format capable of being represented by said feature vectors (page 10, lines 2-3).

Claim 8: The system of claim 1, wherein said documents comprise any of text files, images, web pages, video files, and audio files (page 10, lines 1-2).

Claim 9: The system of claim 2, wherein a classifier at each layer in said hierarchical cascade is trained for each layer with an expectation maximization methodology that maximizes a likelihood of a joint distribution of said training data points and latent variables (page 10, lines 11-13).

Claim 10: The system of claim 9, wherein each layer of said cascade of classifiers is trained in succession from a previous layer by said expectation maximization methodology, wherein said output distribution is used as an input distribution for a succeeding layer (page 10, lines 14-16).

Claim 11: The system of claim 9, wherein each layer of said cascade of classifiers is trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with said output distribution of each layer occurs in succession (page 10, lines 16-19).

Claim 12: The system of claim 11, wherein said successive iterations comprise a fixed number of iterations (page 10, line 19).

Claim 13: The system of claim 9, wherein all layers of said cascade of classifiers are trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with output distributions of all layers occurs, wherein during each step of the of said iterations, the output distribution of each layer is used to weigh the input distribution of a succeeding layer (page 10, line 20 through page 11, line 1).

Claim 14: The system of claim 13, wherein said successive iterations comprise a fixed number of iterations (page 10, line 19).

Claim 15: The system of claim 2, wherein each classifier layer generates a relevancy score associated with each a data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents (page 11, lines 1-3).

Claim 16: The system of claim 2, wherein each classifier layer generates a relevancy score associated with said document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document (page 11, lines 6-9).

Claim 17: The system of claim 5, wherein said terminal classifier generates a relevancy score associated with each data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents, and wherein said relevancy score is computed by combining relevancy scores generated by classifiers at each layer of the cascade (page 11, lines 1-5).

Claim 18: The system of claim 2, wherein said terminal classifier generates a relevancy score associated with a document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document (page 11, lines 6-8).

Claim 19: The system of claim 1, wherein features of said feature vectors comprise words within a range of words located proximate to entities of interest in said document (page 11, lines 16-18).

Claim 20: A method of extracting information, said method comprising: inputting a query (page 9, line 18, Figure 1, block number 100); searching a database of documents based on said query (page 9, lines 18-19, Figure 1, block number 110); retrieving documents from said database based solely on whether said documents are relevant to said query using a plurality of classifiers arranged in a hierarchical cascade of classifier layers (page 9, lines 19-21, Figure 1, block number 120; page 24, line 15 to page 25, line 3), wherein each classifier comprises a set of weighted training data points comprising feature vectors representing any portion of a document (page 9, lines 21-22), wherein each said feature vector is arranged only as a vector of counts for all features in a data point (page 11, lines 12-18); and weighing an output from said cascade according to a rate of success of query terms being matched by each layer of said cascade (page 9, line 22 to page 10, line 1, Figure 1, block number 130), wherein said weighing is performed using a terminal classifier (page 10, line 1).

Claim 21: The method of claim 20, wherein each classifier accepts an input distribution of data points and transforms said input distribution to an output distribution of said data points (page 10, lines 4-6).

Claim 22: The method of claim 21, wherein each classifier is trained by weighing training data points at each classifier layer in said cascade by an output distribution generated by each previous classifier layer, and wherein weights of said training data points of said first classifier layer are uniform (page 10, lines 6-9).

Claim 23: The method of claim 22, wherein each classifier is trained according to said query input (page 10, line 9).

Claim 24: The method of claim 21, wherein said query input is based on a minimum number of example documents (page 10, lines 9-10).

Claim 25: The method of claim 20, wherein said document comprises data points comprising feature vectors representing any portion of said document (page 9, lines 21-22).

Claim 26: The method of claim 20, wherein said documents comprise a file format capable of being represented by said feature vectors (page 10, lines 2-3).

Claim 27: The method of claim 20, wherein said documents comprise any of text files, images, web pages, video files, and audio files (page 10, lines 1-2).

Claim 28: The method of claim 21, wherein a classifier at each layer in said hierarchical cascade is trained for each layer with an expectation maximization methodology that maximizes a

likelihood of a joint distribution of said training data points and latent variables (page 10, lines 11-13).

Claim 29: The method of claim 28, wherein each layer of said cascade of classifiers is trained in succession from a previous layer by said expectation maximization methodology, wherein said output distribution is used as an input distribution for a succeeding layer (page 10, lines 14-16).

Claim 30: The method of claim 28, wherein each layer of said cascade of classifiers is trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with said output distribution of each layer occurs in succession (page 10, lines 16-19).

Claim 31: The method of claim 30, wherein said successive iterations comprise a fixed number of iterations (page 10, line 19).

Claim 32: The method of claim 28, wherein all layers of said cascade of classifiers are trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with output distributions of all layers occurs, wherein during each step of the of said iterations, the output distribution of each layer is used to weigh the input distribution of a succeeding layer (page 10, line 20 through page 11, line 1).

Claim 33: The method of claim 32, wherein said successive iterations comprise a fixed number of iterations (page 10, line 19).

Claim 34: The method of claim 21, wherein each classifier layer generates a relevancy score associated with each a data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents (page 11, lines 1-3).

Claim 35: The method of claim 21, wherein each classifier layer generates a relevancy score associated with said document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document (page 11, lines 6-9).

Claim 36: The method of claim 24, wherein said terminal classifier generates a relevancy score associated with each data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents, and wherein said relevancy score is computed by combining relevancy scores generated by classifiers at each layer of the cascade (page 11, lines 1-5).

Claim 37: The method of claim 21, wherein said terminal classifier generates a relevancy score associated with a document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document (page 11, lines 6-8).

Claim 38: The method of claim 20, wherein features of said feature vectors comprise words within a range of words located proximate to entities of interest in said document (page 11, lines 16-18).

Claim 39: A program storage device readable by computer, tangibly embodying a program of instructions executable by said computer (page 24, lines 3-14, Figure 6) to perform a method of extracting information, said method comprising: inputting a query (page 9, line 18, Figure 1, block number 100); searching a database of documents based on said query (page 9, lines 18-19, Figure 1, block number 110); retrieving documents from said database based solely on whether said documents are relevant to said query using a plurality of classifiers arranged in a hierarchical cascade of classifier layers (page 9, lines 19-21, Figure 1, block number 120; page 24, line 15 to page 25, line 3), wherein each classifier comprises a set of weighted training data points comprising feature vectors representing any portion of a document (page 9, lines 21-22), wherein each said feature vector is arranged only as a vector of counts for all features in a data point (page 11, lines 12-18); and weighing an output from said cascade according to a rate of success of query terms being matched by each layer of said cascade (page 9, line 22 to page 10, line 1, Figure 1, block number 130), wherein said weighing is performed using a terminal classifier (page 10, line 1).

Claim 40: The program storage device of claim 19, wherein each classifier accepts an input distribution of data points and transforms said input distribution to an output distribution of said data points (page 10, lines 4-6).

Claim 41: The program storage device of claim 40, wherein each classifier is trained by weighing training data points at each classifier layer in said cascade by an output distribution generated by each previous classifier layer, and wherein weights of said training data points of said first classifier layer are uniform (page 10, lines 6-9).

Claim 42: The program storage device of claim 41, wherein each classifier is trained according to said query input (page 10, line 9).

Claim 43: The program storage device of claim 40, wherein said query input is based on a minimum number of example documents (page 10, lines 9-10).

Claim 44: The program storage device of claim 39, wherein said document comprises data points comprising feature vectors representing any portion of said document (page 9, lines 21-22).

Claim 45: The program storage device of claim 39, wherein said documents comprise a file format capable of being represented by said feature vectors (page 10, lines 2-3).

Claim 46: The program storage device of claim 39, wherein said documents comprise any of text files, images, web pages, video files, and audio files (page 10, lines 1-2).

Claim 47: The program storage device of claim 40, wherein a classifier at each layer in said hierarchical cascade is trained for each layer with an expectation maximization methodology that maximizes a likelihood of a joint distribution of said training data points and latent variables (page 10, lines 11-13).

Claim 48: The program storage device of claim 47, wherein each layer of said cascade of

classifiers is trained in succession from a previous layer by said expectation maximization methodology, wherein said output distribution is used as an input distribution for a succeeding layer (page 10, lines 14-16).

Claim 49: The program storage device of claim 47, wherein each layer of said cascade of classifiers is trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with said output distribution of each layer occurs in succession (page 10, lines 16-19).

Claim 50: The program storage device of claim 49, wherein said successive iterations comprise a fixed number of iterations (page 10, line 19).

Claim 51: The program storage device of claim 47, wherein all layers of said cascade of classifiers are trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with output distributions of all layers occurs, wherein during each step of the of said iterations, the output distribution of each layer is used to weigh the input distribution of a succeeding layer (page 10, line 20 through page 11, line 1).

Claim 52: The program storage device of claim 51, wherein said successive iterations comprise a fixed number of iterations (page 10, line 19).

Claim 53: The program storage device of claim 40, wherein each classifier layer generates a relevancy score associated with each a data point, wherein said relevancy score comprises an

indication of how closely matched said data point is to said example documents (page 11, lines 1-3).

Claim 54: The program storage device of claim 40, wherein each classifier layer generates a relevancy score associated with said document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document (page 11, lines 6-9).

Claim 55: The program storage device of claim 43, wherein said terminal classifier generates a relevancy score associated with each data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents, and wherein said relevancy score is computed by combining relevancy scores generated by classifiers at each layer of the cascade (page 11, lines 1-5).

Claim 56: The program storage device of claim 40, wherein said terminal classifier generates a relevancy score associated with a document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document (page 11, lines 6-8).

Claim 57: The program storage device of claim 39, wherein features of said feature vectors comprise words within a range of words located proximate to entities of interest in said document (page 11, lines 16-18).

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The issues presented for review by the Board of Patents Appeals and Interferences are

whether claims 1-57 are unpatentable under 35 U.S.C. §103(a) as being unpatentable over Palmer in view of Woo.

VII. ARGUMENT

A. The Prior Art Rejections Based of Claims 1-57

Claims 1-57 are rejected under 35 U.S.C. 103(a) as being unpatentable over Palmer (US Patent No. 6,990,628 B1, filed June 14, 1999) in view of Woo (US Patent No. 7,039,641 B2, filed February 22, 2001).

B. The Position in the Office Action

Regarding claims 1, 20, and 39, according to the Office Action, Palmer discloses a program storage device readable by computer, tangibly embodying a program of instructions executable by said computer to perform a program storage device of extracting information, said program storage device comprising:

inputting a query (Col. 7, lines 30 - 32, Palmer);
searching a database of documents based on said query (Col. 7, lines 28 - 32, Palmer);
retrieving documents from said database based solely on whether said documents are relevant to said query (Col. 4, lines 14- 19; ... the confidence value or score characterizes the relevance of a particular document to a given query documents would be placed in a result set and displaced to a user according to their confidence value or score ...; Palmer) using a plurality of classifiers (Col. 3, lines 1 - 3, Palmer).

The Office Action states that Palmer discloses all the limitations as disclosed above. However, the Office Action admits that Palmer is silent with respect to a hierarchical cascade of

classifier layers. The Office Action then indicates that Woo discloses classifiers arranged in a hierarchical cascade of classifier layers (Fig. 1, item 20, Col. 17, lines 44-49, Woo (wherein the filters correspond to the classifiers claimed)). According to the Office Action it would have been obvious to one of ordinary skill in the art at the time the invention was made to incorporate the Woo' teachings to the system of Palmer. The Office Action further concludes that skilled artisans would have been motivated to do so, as suggested by Woo (Col. 1 and 2, lines 66-67 and 1-2; respectively, Woo), to provide a relatively efficient method and system for finding or identifying an applicable filter when a relatively large number of filters are employed in a packet classification system. Furthermore, the Office Action suggests that both of the references (Palmer and Woo) teach features that are directed to analogous art and they are directed to the same field of endeavor of database management system, such as, searching, classifying data, weights, and frequencies. According to the Office Action, this relation between both of the references highly suggests an expectation of success.

The Office Action states that the combination of Palmer in view of Woo ("Palmer/Woo" hereinafter) further discloses classifiers (Col. 14, lines 23-26, categorization within the training set, Palmer; and Fig. 1, item 20, Col. 17, lines 44-49, Woo (wherein the filters correspond to the classifiers claimed)) including weighted training data points (Col. 13, lines 62-66, Palmer) comprising feature vectors representing any portion of a document (Col. 14, lines 31-35, Palmer (wherein the training set corresponds to the weighted training data points claimed)), wherein each said feature vector is arranged only as a vector of counts for all features in a data point (Col. 14, lines 31-35; each component of the feature vector is the normalized value of the occurrence frequency or a particular feature in this document, Palmer); and weighing an output from said cascade according to a rate of success of query terms being matched by each layer of said

cascade, wherein said weighing is performed using a terminal classifier (Col. 16, lines 1-11, Palmer (wherein the examiner interprets the confidence score as the rate of success claimed; and the β category as the terminal classifier claimed) and Col. 5, lines 54-61, Woo (wherein the examiner interprets $prob(F_i \text{ is the best matching filter for } t)$ as the rate of success claimed)).

Regarding claims 2, 21, and 40, it is the position in the Office Action that Palmer/Woo discloses a program storage device, wherein each classifier accepts an input distribution of data points (Col. 5, lines 41-45, input k-tuple t , Woo) and transforms said input distribution to an output distribution of said data points (Fig. 8, item 72, Col. 5 and 15, lines 41-45 and 20-21, returns the first F_i and output stage; respectively, Woo).

Regarding claims 3, 22, and 41, the Office Action states that Palmer/Woo discloses a program storage device, wherein each classifier is trained by weighing training data points at each classifier layer in said cascade by an output distribution generated by each previous classifier layer (Fig. 8, item 72, Col. 5 and 15, lines 41-45 and 20-21, returns the first F_i and output stage; respectively, Woo), and wherein weights of said training data points of said first classifier layer are uniform (Col. 5, lines 54-57, Woo).

The Office Action then indicates, regarding claims 4, 23, and 42, that Palmer/Woo discloses a program storage device, wherein each classifier is trained according to said query input (Col. 11, lines 46-48, Woo) (wherein the examiner interprets the step of matching the filters to the input packet as the step where each classifier is trained according to the query input claimed and wherein the input bit (further disclosed as “selection criteria” in Col. 11, lines 53-55, Woo) corresponds to the query input claimed).

Regarding claims 5, 24, and 43, the Office Action purports that Palmer/Woo discloses a program storage device, wherein said query input is based on a minimum number of example

documents (Col. 13, lines 62-67, small set of electronic documents, Woo).

Regarding claims 6, 25, and 44, the Office Action indicates that Palmer/Woo discloses a program storage device, wherein said document comprises data points comprising feature vectors representing any portion of said document (Col. 14, lines 31-35, Palmer; and Col. 14, lines 31-35, feature vectors, Woo).

Regarding claims 7, 26, and 45, the Office Action suggests that Palmer/Woo discloses a program storage device, wherein said documents comprise a file format capable of being represented by said feature vectors (Col. 14, lines 31-35, Palmer (the examiner interprets that if feature vectors can be constructed for each document, then it is implied that the format of these documents/or files can be represented in such feature vectors)).

Regarding claims 8, 27, and 46, the Office Action concludes that Palmer/Woo discloses a program storage device, wherein said documents comprise any of text files, images, web pages, video files, and audio files (Col. 3, lines 57-62, Palmer).

Regarding claims 9, 28, and 47, Palmer/Woo discloses a program storage device, wherein a classifier at each layer in said hierarchical cascade is trained for each layer with an expectation maximization methodology that maximizes a likelihood of a joint distribution of said training data points and latent variables (Col. 9, lines 16-21, minimize duplication and maximize “balancedness”, Woo (wherein the distribution of input traffic corresponds to the join distribution claimed)).

Regarding claims 10, 29, and 48, according to the Office Action, Palmer/Woo discloses a program storage device, wherein each layer of said cascade of classifiers is trained in succession from a previous layer by said expectation maximization methodology, wherein said output distribution is used as an input distribution for a succeeding layer (Fig. 8, item 68a, 70a, 68b,

70b, Col. 15, lines 15-19, Woo).

Regarding claims 11, 30, and 49, the Office Action suggests that Palmer/Woo discloses a program storage device, wherein each layer of said cascade of classifiers is trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with said output distribution of each layer occurs in succession (Fig. 4, item 406 and 410, Col. 16, lines 51-55, Palmer (wherein the pre-determined maximum corresponds to convergence parameter of values claimed)).

Regarding claims 12, 31, and 50, the Office Action indicates that Palmer/Woo discloses a program storage device, wherein said successive iterations comprise a fixed number of iterations (Fig. 4, item 406 and 410, Col. 16, lines 51-55, Palmer (the iterations, that include a pre-determined number used for testing, imply a fixed number of iterations as claimed)).

Regarding claims 13, 32, and 51, the Office Action states that Palmer/Woo discloses a program storage device, wherein all layers of said cascade of classifiers are trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with output distributions of all layers occurs (Fig. 4, item 406 and 410, Col. 16, lines 51-55, Palmer (wherein the pre-determined maximum corresponds to convergence parameter of values claimed), wherein during each step of the of said iterations, the output distribution of each layer is used to weigh the input distribution of a succeeding layer (Col. 16, lines 59-62, Palmer (wherein the current value of $x(j, k)$ corresponds to the output distribution claimed; and the buffer $x'(k)$ corresponds to the input distribution claimed) and Fig. 8, item 68a, 70a, 68b, 70b, Col. 15, lines 15-19, Woo).

Regarding claims 14, 33, and 52, the Office Action suggests that Palmer/Woo discloses a program storage device, wherein said successive iterations comprise a fixed number of iterations

(Fig. 4, item 406 and 410, Col. 16, lines 51-55, Palmer (the iterations, that include a pre-determined number used for testing, imply a fixed number of iterations)).

Regarding claims 15, 35, and 53, it is the position of the Office Action that Palmer/Woo discloses a program storage device, wherein each classifier layer generates a relevancy score associated with each a data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents (Col. 4, lines 11-16, Palmer).

Regarding claims 16, 37, and 54, the Office Action indicates that Palmer/Woo discloses a program storage device, wherein each classifier layer generates a relevancy score associated with said document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document (Col. 4, lines 14-16, confidence value or score characterizes the relevance of a particular document to a given query, Palmer).

Regarding claims 17, 36, and 55, the Office Action posits that Palmer/Woo discloses a program storage device, wherein said terminal classifier generates a relevancy score associated with each data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents (Col. 4, lines 11-16, Palmer), and wherein said relevancy score is computed by combining relevancy scores generated by classifiers at each layer of the cascade (Col. 12, lines 22-28, Palmer).

Regarding claims 18, 34, and 56, the Office Action indicates that Palmer/Woo discloses a program storage device, wherein said terminal classifier generates a relevancy score associated with a document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document (Col. 4, lines 14-16, Palmer).

Regarding claims 19, 38, and 57, the Office Action concludes that Palmer/Woo discloses a program storage device, wherein features of said feature vectors comprise words within a range

of words located proximate to entities of interest in said document (Col. 14, lines 29-35, Palmer (the examiner interprets one-half million word phrases out of two million features corresponds to the words within a range claimed)).

C. The Prior Art References

1. The Palmer Reference

Palmer teaches a method and apparatus for determining when electronic documents stored in a large collection of documents are similar to one another. A plurality of similarity information is derived from the documents. The similarity information may be based on a variety of factors, including hyperlinks in the documents, text similarity, user click-through information, similarity in the titles of the documents or their location identifiers, and patterns of user viewing. The similarity information is fed to a combination function that synthesizes the various measures of similarity information into combined similarity information. Using the combined similarity information, an objective function is iteratively maximized in order to yield a generalized similarity value that expresses the similarity of particular pairs of documents. In an embodiment, the generalized similarity value is used to determine the proper category, among a taxonomy of categories in an index, cache or search system, into which certain documents belong.

2. The Woo Reference

Woo teaches a method and system for classifying packets through the use of filters and combines heuristic tree searches with the use of filter buckets. This provides high performance and reasonable storage requirements, even when applied to large number of filters (from 4K to 1

million). In addition, the method can adapt to the input packet distribution by taking into account the relative filter usage. The capability of employing a large number of filters in a packet classification system is useful in providing value-added services, such as security, quality of service (QoS), load balancing, and traffic accounting.

D. The Appellants' Position

1. Independent claims 1, 20, and 39

The Office Action cites Palmer and Woo as teaching the Appellants' claimed invention as defined by claims 1, 20, and 30. However, the Appellants respectfully but strongly disagree with this conclusion. In particular, the Appellants disagree that in FIG. 1 of Palmer, various types of information such as hyperlink information, text similarity, multimedia component similarity, URL similarity, click-through information, and cache hit log similarity are combined together to create a similarity objective function, which is then used to compare two documents in order to create generalized similarity values among the two documents. Thus, in Palmer the pages are retrieved first and then associated a similarity value and then, as indicated in col. 11, lines 36-40 of Palmer, "the generalized similarity value 120 is used to determine the proper category, among a taxonomy of categories in an index, cache or search system, into which documents 112, 114 belong." This indicates that the similarity value is used for classification purposes; i.e., how to group the documents. Conversely, in the Appellants' claimed invention, no such classification takes place; rather documents are retrieved based solely on whether they are relevant to the query input. In Palmer, FIG. 1 and the associated text indicates that documents are retrieved based on many factors, some not necessarily related to relevancy to a particular query input, and are compared with other documents to determine a classification of

the documents. Moreover, col. 5, lines 41-45 of Palmer teach finding dissimilar documents, “to find dissimilar documents, only the documents in the most dissimilar categories are compared. As a result, the most similar and the most dissimilar pairs of documents tend to be obtained. For these pairs, positively weighted or negatively weighted text links are created and stored.” This teaches away from the Appellants’ claimed “classifiers are operable to retrieve documents from said database based solely on whether said documents are relevant to said query input.”
Accordingly, Palmer teaches away from the Appellants’ claimed invention.

Additionally, Palmer deals with similarities between documents, using specific features such as URLs. Palmer uses a training algorithm to iteratively maximize an objective function of the similarity matrix. Therefore, Palmer requires that the amount of calculation and storage be proportional to the square of the number of documents. Conversely, the Appellants’ invention uses feature vectors only (i.e., solely) instead of matrices, which require only resources to be proportional to the number of documents. Moreover, Palmer’s classification of documents is performed independent of a specific query. For example, in Palmer the classification occurs, in part, based on the time spent by a user browsing various documents (column 10, lines 15-34). Conversely, the Appellants’ invention only performs its classification to determine whether documents are relevant to a specific query, ignoring them in all other aspects not relevant to the query. This is analogous to the Appellants claims 1, 20, and 39, which respectively state, “wherein said classifiers are operable to retrieve documents from said database based solely on whether said documents are relevant to said query input” (claim 1) and “retrieving documents from said database based solely on whether said documents are relevant to said query using a plurality of classifiers...” (claims 20 and 39).

Additionally, Palmer calculates a weight matrix describing similarity information $w(i,j)$

for each pair of documents i and j (column 12, lines 18-19). This is calculated from a number of relationships between these two documents (column 4, line 27 to column 10, line 51). This calculation is performed beforehand (i.e., prior to query input), independent of user queries (column 2, lines 64-68, and column 3, lines 1-3). In contrast, the Appellants' invention only uses the rate of success of relevance (i.e., matching) of each layer of the cascade to the specific topic of the user query (i.e., after the query input). Accordingly, the Appellants' invention does not have to calculate the similarity between each pair of documents. In other words, Palmer's invention does not take into account the user query, and it requires the availability, calculation, and storage of pairwise information, which is not applicable in the Appellants' invention. Conversely, the Appellants' invention does not make use of pairwise information even when it is available.

The work of Woo deals with packet filters for network traffic where a large number of simple filters are employed to process a large number of packets at a very fast rate. These filters perform very simple binary decisions. The emphasis in Woo is on getting through as many filters as possible by bypassing branches of the decision tree. The applicability of Woo's invention relies on the fact each filter can make deterministic binary decisions. Woo also talks about weights (column 5, lines 53-67), however the values calculated from these weights are again used to shape the decision tree (columns 11 and 12); i.e., bypassing branches on the decision tree.

In contrast, the Appellants' claimed invention propagates probability (feature) vectors from one classifier (filter) to another (i.e., "each layer of the cascade of classifiers is trained in succession from a previous layer by the expectation maximization methodology, wherein the output distribution is used as an input distribution for a succeeding layer."). Moreover, in Woo,

results are thresholded to zero or one (i.e., according to claim 5 of Woo “each non-leaf node has two child nodes, one said child node representing a set of filters having a 0 or * bit at the bit position corresponding to the non-leaf node, and the other said child node representing a set of filters having a 1 or * bit at the bit position corresponding to the non-leaf node.”). Conversely, in the Appellants’ invention, results calculated from each classifier layer are not thresholded to zero or one. Accordingly, Woo teaches away from the Appellants’ claimed invention.

Furthermore, one of the applications for the Appellants’ invention is in text analytics, where it is known to those skilled in the art that the thresholding of the results or bypassing of the layers most often degrades the quality of the result. Accordingly, the Appellants’ claimed invention is patentably distinct from Woo (in combination with Palmer).

The work of Woo is in a completely different field (i.e., packet filtering in network traffic) from either the Appellants’ invention or Palmer’s invention (i.e., text information extraction). On the one hand, in Woo the packets arrive in an endless time sequence, and have to be disposed of quickly, independent of each other (column 1, lines 38-47). Neither Palmer’s invention nor the Appellants’ invention are applicable to this field, since both require an iterative calculation on a fixed large collection of documents. On the other hand, Woo’s invention assumes that the filter rules are all fixed beforehand (column 17, lines 44 - 46). Therefore it is not applicable to text information retrieval, where it is important to learn the filter rules (weight matrix in Palmer’s invention and feature vectors in the Appellants’ invention). Woo’s invention bypasses a large number of filters yet still provides an answer as if all the filters have been consulted (column 1, line 66 to column 2, line 2). This is only possible if all the filters make yes-no (i.e., zero or one) decisions; conversely in the Appellants’ invention the classifiers in text information retrieval inherently keep weights or probabilities which are not zero or one. Since

Woo's invention solves problems completely different from the problems solved by either Palmer or the Appellants, the above contrast is made only approximately, assuming that packets correspond to documents and filters correspond to classifiers, as implicated by the last paragraph of page 3 of the Office Action. In reality, packets are very simple and well-defined entities (see Woo, column 1, lines 26-32), while texts are complex and unstructured (see Palmer, column 4, line 27 to column 10, line 51). Accordingly, Woo in combination with Palmer would result in an inoperable device/method producing conflicting results.

Furthermore, the comment on pages 6-8 of the Office Action suggests that Woo's invention discloses an "expectation maximization methodology that maximizes a likelihood of a joint distribution of said training data points and latent variables". The term "expectation maximization methodology" in the Appellants' invention refers to a specific statistical procedure that is applicable to systems with a specific kind of statistical models (a latent variable model is one of them). There is nothing in Woo (column 9, lines 16-21) that remotely relates to any of this. The following terms are technical terminology related to statistical models: expectation maximization algorithm (or methodology), likelihood, joint distribution, latent variables. However, these terms do not apply to Woo's setting where a statistical model is absent. There is no basis in either Woo or in any other prior art reference or by the vernacular used those skilled in the art that uses "minimize duplication" and "maximize balancedness" to be analogous to an "expectation maximization methodology". In fact, the Appellants are well versed in the work done by their contemporaries, have attended conferences, have read articles, have researched others' work, and have published their own articles and have never heard such use of the language as is being suggested in the Office Action to refer to the Appellants' "expectation maximization methodology". Accordingly, Woo cannot possibly teach the Appellants' claimed

invention.

Moreover, the comment on page 4 of the Office Action about “the same field of endeavor of database management system, such as, searching, classifying data, weights and frequencies” suggests a possible misunderstanding of the terminology. Database systems store data and allow the user to change and search/retrieve them later. However, those skilled in the art would readily acknowledge that a packet filtering system, such as in Woo, makes decisions on each packet and promptly forgets about them. Searching is only possible if data is stored. The “frequency” in Woo is the physical frequency, which refers to how many packets arrive in a period of time, while the “frequency” as used by the Appellants’ is the statistical frequency, such as how many times a word appears in a document.

In Woo, frequency is an operational requirement, while in the Appellants’ invention, frequency is one of the features used in the calculation. In a non-technical sense, all three systems can be loosely said to “classify data”. However, the Federal Circuit cautions not to read claims in a vacuum, but rather in the context of the specification. *In re Marosi*, 710 F.2d 799, 802 218 USPQ 289, 292 (Fed. Cir. 1983) (quoting *In re Okuzawa*, 537 F.2d 545, 548, 190 USPQ 464, 466 (CCPA 1976)). Thus, the Appellants’ claims must be in light of the language in the specification, and in doing so illustrates that the Appellants’ claims are (a) novel over the prior art, (b) unobvious over the prior art, and (c) non-analogous to the prior art.

Pages 11-12 of the Office Action (and reiterated in the Advisory Action) states that “Applicant argues that the prior art fails to disclose ‘the concept of collection’” and that ““the concept of collection’ is not recited in the rejected claim(s).” However, the Appellants have never stated that they are claiming “the concept of collection.” Rather, this is being offered to prove that Woo and Palmer are non-analogous art. As indicated, generally, Woo makes a quick

decision on packets as they arrive, with predefined rules; Palmer classifies all documents into categories and calculates similarities to be used later; while the Appellants' invention classifies documents relative to a particular user query (and solely based on this). The result of the classification in both Palmer and the Appellants' invention depends on what other documents are in the collection, while the concept of collection is absent from Woo. Because Woo fails to teach this, while Palmer does, it is highly indicative that Palmer and Woo are non-analogous. Woo uses weights not to make decisions, but rather to choose decision makers (filters) so as to save time; the final decision should not depend on the weights if the invention in Woo is functionally correct and operable. Accordingly, the Appellants' invention provides a manner of weighing an output from the cascade of classifier layers according to a rate of success of query terms being matched by each layer of the cascade of classifier layers. Neither Palmer nor Woo teaches this.

Insofar as references may be combined to teach a particular invention, and the proposed combination of Palmer with Woo, case law establishes that, before any prior-art references may be validly combined for use in a prior-art 35 U.S.C. § 103(a) rejection, the individual references themselves or corresponding prior art must suggest that they be combined.

For example, in In re Sernaker, 217 USPQ 1, 6 (C.A.F.C. 1983), the court stated: “[P]rior art references in combination do not make an invention obvious unless something in the prior art references would suggest the advantage to be derived from combining their teachings.” Furthermore, the court in Uniroyal, Inc. v. Rudkin-Wiley Corp., 5 USPQ 2d 1434 (C.A.F.C. 1988), stated, “[w]here prior-art references require selective combination by the court to render obvious a subsequent invention, there must be some reason for the combination other than the hindsight gleaned from the invention itself. . . . Something in the prior art must suggest the

desirability and thus the obviousness of making the combination.”

In the present application, the reason given to support the proposed combination is improper, and is not sufficient to selectively and gratuitously substitute parts of one reference for a part of another reference in order to try to meet, but failing nonetheless, the Appellant’s novel claimed invention. Furthermore, the claimed invention, as amended, meets the above-cited tests for obviousness by including embodiments such as “retrieving documents from said database based solely on whether said documents are relevant to said query” and “wherein each said feature vector is arranged only as a vector of counts for all features in a data point.” As such, all of the claims of this application are, therefore, clearly in condition for allowance, and it is respectfully requested that the Examiner pass these claims to allowance and issue.

As declared by the Federal Circuit:

In proceedings before the U.S. Patent and Trademark Office, the Examiner bears the burden of establishing a prima facie case of obviousness based upon the prior art. The Examiner can satisfy this burden only by showing some objective teaching in the prior art or that knowledge generally available to one of ordinary skill in the art would lead that individual to combine the relevant teachings of the references. In re Fritch, 23 USPQ 2d 1780, 1783 (Fed. Cir. 1992) citing In re Fine, 5 USPQ 2d 1596, 1598 (Fed. Cir. 1988).

Here, the Examiner has not met the burden of establishing a prima facie case of obviousness. Page 13 of the Office Action suggests that the Examiner has met this burden. However, the Appellants respectfully disagree. Page 13 of the Office Action indicates that col. 1 and 2, liens 63-67, and 1-2, respectively of Woo provide the suggestion or motivation to combine the teachings of Woo with Palmer. However, the aforementioned passage in Woo merely states, “A more pragmatic approach is desired. In particular, it is [desirable] to be able to classify packets using a relatively large number of filters given the present state of packet arrival rates.

Towards this end the invention seeks to provide a relatively efficient method and system for finding or identifying an applicable filter when a relatively large number of filters are employed in a packet [classification] system.” As previously, discussed Woo deals with packet filtering in network traffic while Palmer deals with text information extraction. Accordingly, Woo does not deal with extracting text information and there is nothing in the above language in Woo that suggests its teachings are meant to be combined with a system or method of text information extraction. Any reading of this into the above passage of Woo is unreasonably broad and is not something that one of ordinary skill in the art would partake in. Next, the Office Action states that Woo suggests a successful outcome of the combination [with Palmer]. Again, Woo seeks to provide a relatively efficient method and system for finding or identifying an applicable filter when a relatively large number of filters are employed in a packet classification system. Since Palmer does not deal with packet filtering, there is no reasonable basis to conclude that Woo is seeking to identify an applicable filter in a text extraction system or method. Next, the Office Action states that Palmer and Woo teach features that are directed to the same industry field. However, while some features may overlap, the totality of the subject matter disclosed in Palmer and Woo are from separate industry fields.

In fact, the USPTO in classifying Woo and Palmer has essentially determined that they are from non-analogous art fields. For example, the USPTO has classified Palmer in U.S. Classes 715/500; 715/501.1; 707/3; and 707/6. Conversely, the USPTO has classified Woo in U.S. Classes 707/100; 370/392; 370/389; and 370/401. Thus, one of ordinary skill in the art would not have been motivated to combine Palmer with Woo especially considering that the USPTO, an expert organization having a breadth of resources at its availability, itself offers no suggestion of such a combination.

Accordingly, it is clear that, not only does Palmer fail to disclose all of the elements of the claims of the present invention, particularly, “retrieving documents from said database based solely on whether said documents are relevant to said query” and “wherein each said feature vector is arranged only as a vector of counts for all features in a data point,” as discussed above, but also, if combined with Woo, fails to disclose these elements as well. The unique elements of the claimed invention are clearly an advance over the prior art.

The Federal Circuit also went on to state:

The mere fact that the prior art may be modified in the manner suggested by the Examiner does not make the modification obvious unless the prior art suggested the desirability of the modification. . . . Here the Examiner relied upon hindsight to arrive at the determination of obviousness. It is impermissible to use the claimed invention as an instruction manual or “template” to piece together the teachings of the prior art so that the claimed invention is rendered obvious. This court has previously stated that one cannot use hindsight reconstruction to pick and choose among isolated disclosures in the prior art to deprecate the claimed invention. Fritch at 1784-85, citing In re Gordon, 221 USPQ 1125, 1127 (Fed. Cir. 1984).

Here, there is no suggestion that Palmer, alone or in combination with Woo teaches a method and structure containing all of the limitations of the claimed invention. Consequently, there is absent the “suggestion” or “objective teaching” that would have to be made before there could be established the legally requisite “prima facie case of obviousness.”

Additionally, clearly the Appellants’ claimed invention is part of a crowded art field as evidenced by the references cited in the Appellants’ specification and provided in the Information Disclosure Statement filed by the Appellants on November 26, 2003 as well as the multiple references considered as pertinent prior art cited in the Office Action of May 19, 2006, and furthermore the several references cited in each of the prior art references cited by the Office

Action. As such, given the crowdedness of the art, the novel aspects of the Appellants' claimed invention should be regarded as a significant step forward in the constant development of this technical art field.

Additionally, much of the Office Action's rejections are based on the Examiner's stretched interpretation of the prior art and attempting to establish a link to some of the features in the Appellants' claims. However, the totality of the claimed features must be analyzed and taught in the prior art for a proper rejection under 35 USC §103(a). It appears that the Examiner is not reading the claims in their totality (i.e., as a whole), but rather is arbitrarily selecting portions of the claims and attempting to create a link to portions of the prior art in a piecemeal manner. This conflicts with MPEP §§2116.01; 2141; 2141.02; and 2142 and 35 USC§103(a), and as such, the rejection is improper.

In view of the foregoing, the Appellants respectfully submit that Palmer and Woo do not teach or suggest the features defined by independent claims 1, 20, and 39 and as such, claims 1, 20, and 39 are patentable over Palmer in combination with Woo. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to independent claims 1, 20, and 39.

2. Dependent claims 2, 21, and 40

The comments on page 5 of the Office Action (and reiterated in the Advisory Action) regarding Appellants' claims 2, 21, 40 suggests that Woo's invention deals with "input distribution" and "output distribution", but there are no such concepts in Woo in the columns/lines cited in the Office Action. Instead, Woo only deals with "input tuples" and "output tuples", which those skilled in the art would readily understand are binary numbers. In

contrast, a “distribution” in the context of the Appellants’ invention is not a single tuple, but an assignment of probabilities to all possible such tuples (see Appellants’ specification, page 13, lines 14-17; page 14, lines 9-21; page 15, lines 13-19). Accordingly, as the MPEP suggests, the Appellants may be their own lexicographers (MPEP §2111.01(III)), and as such the Appellants’ claimed language should be read in light of the definitions and descriptions provided in the Appellants’ specification. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants’ claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 2, 21, and 40.

3. Dependent claims 3, 22, and 41

The comments on page 5 of the Office Action (and reiterated in the Advisory Action) regarding Appellants’ claims 3, 22, and 41 suggests that Woo’s invention deals with “input distribution” and “output distribution”, but there are no such concepts in Woo in the columns/lines cited in the Office Action. Instead, Woo only deals with “input tuples” and “output tuples”, which those skilled in the art would readily understand are binary numbers. In contrast, a “distribution” in the context of the Appellants’ invention is not a single tuple, but an assignment of probabilities to all possible such tuples (see Appellants’ specification, page 13, lines 14-17; page 14, lines 9-21; page 15, lines 13-19). Accordingly, as the MPEP suggests, the Appellants may be their own lexicographers (MPEP §2111.01(III)), and as such the Appellants’ claimed language should be read in light of the definitions and descriptions provided in the Appellants’ specification.

Furthermore, the Office Action cites Col. 5, lines 54-57 of Woo as teaching the Appellants’ “wherein weights of said training data points of said first classifier layer are

uniform.” However, this portion of Woo merely states, “[t]he weight represents the relative match frequency of a particular filter, and is typically derived from the distribution of the input tuple t or filter usage statistics.” Clearly there is nothing in the cited portion of Woo remotely suggesting the Appellants’ claimed invention pertaining to weights of training data points of a first classifier layer being uniform. As such, it is unclear how the Office Action is interpreting Woo in this regard in order to reject the Appellants’ claims. Therefore, the Appellants submit that this rejection is deficient as citing prior art not teaching the Appellants’ claimed invention as defined by claims 3, 22, and 41. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants’ claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 3, 22, and 41.

4. Dependent claims 4, 23, and 42

The Office Action cites Col. 11, lines 46-48 and 53-55 of Woo as teaching the Appellants’ claimed invention defined by dependent claims 4, 23, and 42. The relevant portion of Woo teaches, “[t]he rationale is that if the b -th bit of an input packet is “0,” then it can only match the filters in the “0”-group and thus only those need to be considered further, and vice versa for the “1”-group...Those skilled in this art will appreciate, however, that the bit selection criteria described below extends in a straightforward manner to the multibit case.” The Office Action indicates that the Examiner interprets this language as teaching “wherein each classifier is trained according to said query input.” The Appellants contend that this interpretation is improperly drawn because while the Examiner interprets the input bit as a selection criteria as corresponding to the Appellants’ query input, a closer review of Woo’s teaching of its bit selection criteria given in Col. 11, line 62 through Col. 12, line 24, which provides, in part:

2.4 Bit Selection

The bit selected at each node determines the overall “shape” of the tree. Thus, given some global measure of the “goodness” of a search tree, the bit selected at each node should ideally “grow” the tree toward some final optimal shape. In abstract terms, we assign a preference value for each unprocessed bit position (step (2.3)), and we pick the bit with the highest preference position (step (2.4)).

For a search tree, a typical “goodness” measure is the weighted average search path length...This measure, though concrete and optimal, is computationally expensive to calculate, as it involves comparing fully constructed trees.

As a compromise, the present approach tries to optimize local measures in a hope that they cumulatively produce a reasonably “good” global solution. The “localness” of a measure is defined by the amount of lookaheads it uses. In what follows, results are presented only for the case where a single bit is chosen at each node and the preference value is based only on one level of lookahead.

The above language of Woo suggests that its preferred implementation of its input bits are not used to train each classifier, but rather uses a single bit assigned at each node, which corresponds to a preference value that is only one level removed (lookahead) from the current node. Furthermore, the above language of Woo also suggests selectively determining the number of “lookaheads” to capture a certain level of “localness”, which is further evidence that each classifier is not trained according to the query input in Woo because one could select certain classifiers as not being “local” and thereby not being trained. Conversely, the Appellants’ classifiers are arranged in a hierarchical cascade of classifiers (i.e., not limited to merely one level of lookahead) and as such the suggested interpretation in the Office Action would render Woo unsatisfactory for its intended purpose, which according to MPEP §2143.01, renders the current obviousness rejection improper. Furthermore, even a combination of Palmer with Woo

fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 4, 23, and 42.

5. Dependent claims 5, 24, and 43

The Office Action suggests that Col. 13, lines 62-67 Woo teach a "small set of electronic documents." However, this language never appears in the cited section of Woo, and in fact, a global search of Woo reveals no such language appearing anywhere in Woo. Clearly, this language is merely the Examiner's own personal interpretation of Woo and without tangible evidence establishing that Woo teaches or reasonably suggests that a query input is based on a minimum number of example documents or that one of ordinary skill in the art would interpret such a feature in Woo, this interpretation is improperly drawn rendering the obviousness rejection of claims 5, 24, and 43 defective. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 5, 24, and 43.

6. Dependent claims 6, 25, and 44

The Office Action indicates that Col. 14, lines 31-35 of Palmer teaches the Appellants' "wherein said document comprises data points comprising feature vectors representing any portion of said document." However, the cited portion of Palmer merely states, "[a]s shown in block 316, word vectors or feature vectors are constructed for each document in the expanded set. Each component of the feature vectors is the normalized value of the occurrence frequency

of a particular feature in this document.” This language of Palmer merely suggests that feature vectors are constructed for each document. It does not indicate that the feature vectors represent any portion of the document. In fact, a different teaching is suggested in the above language of Palmer by indicating that the feature vectors represent the frequency of occurrence of a particular feature in a document (i.e., a discrete number rather than a portion of document). The Office Action goes on to suggest that Col. 14, liens 31-35 of Woo teach the Appellants’ claimed language. However, closer review of the cited portion of Woo reveals that Woo does not even refer to feature vectors in this cited portion of Woo or in the entire teaching of Woo. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants’ claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 6, 25, and 44.

7. Dependent claims 7, 26, and 45

The Office Action indicates that Col. 14, lines 31-35 of Palmer teaches the Appellants’ “wherein said documents comprise a file format capable of being represented by said feature vectors.” However, the cited portion of Palmer merely states, “[a]s shown in block 316, word vectors or feature vectors are constructed for each document in the expanded set. Each component of the feature vectors is the normalized value of the occurrence frequency of a particular feature in this document.” This language of Palmer merely suggests that feature vectors are constructed for each document. It does not indicate what the characteristics are of the document, and any reading of this in Palmer is overly restrictive and would not be obvious to one of ordinary skill in the art due to the level of one of ordinary skill in the art being a computer programmer. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of

the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 7, 26, and 45.

8. Dependent claims 8, 27, and 46

The Office Action refers to Col. 3, lines 57-62 of Palmer as teaching the Appellants' "wherein said documents comprise any of text files, images, web pages, video files, and audio files." However, the cited portion of Palmer merely recites, "a Similarity may derive from similarity in document text; a word vector may describe the distribution of word occurrences in a document, and the dot product of two word vectors may characterize the text similarity of two documents. A Similarity may also include equivalent images, audio components, or other multimedia elements." This language of Palmer clearly indicates that the Similarity (which Palmer defines as "a Similarity among electronic documents is defined broadly to mean a relation between pages, including but not limited to hypertext links" as provided in Col. 3, lines 52-54 in Palmer) may include images, audio components, or other multimedia elements, but it does not teach or suggest that the document itself includes images, audio components, or other multimedia elements. Therefore, Palmer is deficient in teaching all the elements of the Appellants' claimed invention defined by claims 8, 27, and 46. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 8, 27, and 46.

9. Dependent claims 9, 28, and 47

The Office Action indicates that Col. 9, lines 16-21 of Woo teaches the Appellants'

claimed invention defined by claims 9, 28, and 47. The cited portion of Woo teaches “the choice is made to minimize duplication and maximize “balancedness” of the 2ⁿ children. Many different criteria can be defined for the division. An embodiment presented below takes into account the filter usage statistics, thus allowing it to adapt to the distribution of input traffic.” The Office Action interprets this as teaching, “wherein a classifier at each layer in said hierarchical cascade is trained for each layer with an expectation maximization methodology that maximizes a likelihood of a joint distribution of said training data points and latent variables.” However, it appears that the Examiner is not reading the claims in their totality (i.e., as a whole), but rather is arbitrarily selecting portions of the claims and attempting to create a link to portions of the prior art in a piecemeal manner. This conflicts with MPEP §§2116.01; 2141; 2141.02; and 2142 and 35 USC§103(a), and as such, the rejection is improper. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants’ claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 9, 28, and 47.

10. Dependent claims 10, 29, and 48

The Office Action indicates that Fig. 8, item 68a, 70a, 68b, 70b, and Col. 15, line 15-19 of Woo teach the Appellants’ “wherein each layer of said cascade of classifiers is trained in succession from a previous layer by said expectation maximization methodology, wherein said output distribution is used as an input distribution for a succeeding layer.” However, the cited sections of Woo reveal no such teaching. Instead, the corresponding section of Woo teaches, “[b]ucket search 68, which has its own internal pipeline. Each stage 68a, 68b, . . . , 68n of this M-stage internal pipeline handles BUCKETDEPTH/M filters. For each stage to operate in

parallel, each has its own memory bank 70a, 70b, . . . , 70n for storing the selected section of the filters. Output stage 72, which retrieves action data 74 corresponding to the match.” There is nothing in this language which refers to or even suggests an expectation maximization methodology or its reasonable equivalence, let alone using an expectation maximization methodology to train each layer of a cascade of classifiers or filters. Rather, Woo merely teaches a bank of memory components having input and output stages. Moreover, Fig. 8 of Woo is also silent as to the Appellants’ claimed features described in claims 10, 29, and 48. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants’ claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 10, 29, and 48.

11. Dependent claims 11, 30, and 49

The Office Action indicates that Fig. 4, item 406 and 410 and Col. 16, lines 51-55 of Palmer teach the elements of the Appellants’ claims 11, 30, and 49. Palmer recites, “[i]n one embodiment, the classification confidence vectors are represented as $X(j, k)=1$ if page k is a training document that belongs to category j; otherwise the vector element is set equal to In block 406, a loop is entered to carry out training iterations. A counter variable i is initialized to “0”. Block 406 also involves testing whether the value of i is less than a pre-determined maximum. If not, then training is complete, and the output is written to storage, as indicated by block 408.” As indicated in the above portion of Palmer, there is no teaching in Palmer of the Appellants’ “wherein each layer of said cascade of classifiers is trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with said output distribution of each layer occurs in succession,” especially considering that page 3 of

the Office Action admits that Palmer does not teach a cascade of classifier layers. Thus, the cited portions of Palmer including Fig. 4 cannot teach the entirety of the Appellants' claims 11, 30, and 49. It appears that the Examiner is not reading the claims in their totality (i.e., as a whole), but rather is arbitrarily selecting portions of the claims and attempting to create a link to portions of the prior art in a piecemeal manner. This conflicts with MPEP §§2116.01; 2141; 2141.02; and 2142 and 35 USC§103(a), and as such, the rejection is improper. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 11, 30, and 49.

12. Dependent claims 12, 31, and 50

The Office Action cites Fig. 4, item 406 and 410 and Col. 16, lines 51-55 as teaching the Appellants' "said successive iterations comprise a fixed number of iterations." Palmer recites, "[i]n one embodiment, the classification confidence vectors are represented as $X(j, k)=1$ if page k is a training document that belongs to category j ; otherwise the vector element is set equal to In block 406, a loop is entered to carry out training iterations. A counter variable i is initialized to "0". Block 406 also involves testing whether the value of i is less than a pre-determined maximum. If not, then training is complete, and the output is written to storage, as indicated by block 408." As indicated in the above language of Palmer, there is no teaching in Palmer of the Appellants' claimed elements provided in claims 12, 31, and 50. Rather, as indicated above, Palmer merely refers to testing whether a value of i is less than a pre-determined maximum. However, there is no such testing in the Appellants' claims, nor is there a counter variable. Furthermore, the Appellants' iterations are non-analogous to Palmer's counter variable because

Palmer's counter variable has nothing to do with an expectation maximum methodology, whereas the Appellants' iterations are iterations of an expectation maximum methodology, which is described in claims 11, 30, and 49, which claims 12, 31, and 50 respectively depend from, and as such the limitations of claims 11, 30, and 49 should be read into claim 12, 31, and 50, respectively, for purposes of patentability. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 12, 31, and 50.

13. Dependent claims 13, 32, and 51

The Office Action cites Fig. 4, item 406 and 410 and Col. 16, lines 51-55 as teaching the Appellants' "wherein all layers of said cascade of classifiers are trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with output distributions of all layers occurs, wherein during each step of the of said iterations, the output distribution of each layer is used to weigh the input distribution of a succeeding layer." Palmer recites, "[i]n one embodiment, the classification confidence vectors are represented as $X(j, k)=1$ if page k is a training document that belongs to category j; otherwise the vector element is set equal to In block 406, a loop is entered to carry out training iterations. A counter variable i is initialized to "0". Block 406 also involves testing whether the value of i is less than a pre-determined maximum. If not, then training is complete, and the output is written to storage, as indicated by block 408." As indicated in the above language of Palmer, there is no teaching in Palmer of the Appellants' claimed elements provided in claims 13, 32, and 51. Rather, as indicated above, Palmer merely refers to testing whether a value of i is less than a pre-

determined maximum. However, there is no such testing in the Appellants' claims, nor is there a counter variable. Furthermore, the Appellants' iterations are non-analogous to Palmer's counter variable because Palmer's counter variable has nothing to do with an expectation maximum methodology, whereas the Appellants' iterations are iterations of an expectation maximum methodology. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 13, 32, and 51.

14. Dependent claims 14, 33, and 52

The Office Action cites Fig. 4, item 406 and 410 and Col. 16, lines 51-55 as teaching the Appellants' "said successive iterations comprise a fixed number of iterations." Palmer recites, "[i]n one embodiment, the classification confidence vectors are represented as $X(j, k)=1$ if page k is a training document that belongs to category j ; otherwise the vector element is set equal to In block 406, a loop is entered to carry out training iterations. A counter variable i is initialized to "0". Block 406 also involves testing whether the value of i is less than a pre-determined maximum. If not, then training is complete, and the output is written to storage, as indicated by block 408." As indicated in the above language of Palmer, there is no teaching in Palmer of the Appellants' claimed elements provided in claims 14, 33, and 52. Rather, as indicated above, Palmer merely refers to testing whether a value of i is less than a pre-determined maximum. However, there is no such testing in the Appellants' claims, nor is there a counter variable. Furthermore, the Appellants' iterations are non-analogous to Palmer's counter variable because Palmer's counter variable has nothing to do with an expectation maximum methodology, whereas the Appellants' iterations are iterations of an expectation maximum methodology, which

is described in claims 13, 32, and 51, which claims 14, 33, and 52 respectively depend from, and as such the limitations of claims 13, 32, and 51 should be read into claim 14, 33, and 52, respectively, for purposes of patentability. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 14, 33, and 52.

15. Dependent claims 15, 34, and 53

The Office Action refers to Col. 4, lines 11-16 of Palmer as teaching the elements of the Appellants' claims 15, 34, and 53. In fact, the above cited section of Palmer merely states, “[t]he sum of the distances for a document is a confidence value or score. In a directory engine, the confidence value or score characterizes the confidence that a particular document falls within a particular category. In a search engine, the confidence value or score characterizes the relevance of a particular document to a given query.” However, there is nothing in the above language of Palmer that teaches or reasonably suggests that each classifier or filter layer generates the relevancy score, which the Appellants' provide. It appears that the Examiner is not reading the claims in their totality (i.e., as a whole), but rather is arbitrarily selecting portions of the claims and attempting to create a link to portions of the prior art in a piecemeal manner. This conflicts with MPEP §§2116.01; 2141; 2141.02; and 2142 and 35 USC§103(a), and as such, the rejection is improper. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 15, 34, and 53.

16. Dependent claims 16, 35, and 54

The Office Action refers to Col. 4, lines 11-16 of Palmer as teaching the elements of the Appellants' claims 16, 35, and 54. In fact, the above cited section of Palmer merely states, “[t]he sum of the distances for a document is a confidence value or score. In a directory engine, the confidence value or score characterizes the confidence that a particular document falls within a particular category. In a search engine, the confidence value or score characterizes the relevance of a particular document to a given query.” However, there is nothing in the above language of Palmer that teaches or reasonably suggests that each classifier or filter layer generates the relevancy score, which the Appellants' provide. It appears that the Examiner is not reading the claims in their totality (i.e., as a whole), but rather is arbitrarily selecting portions of the claims and attempting to create a link to portions of the prior art in a piecemeal manner. This conflicts with MPEP §§2116.01; 2141; 2141.02; and 2142 and 35 USC §103(a), and as such, the rejection is improper. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 16, 35, and 54.

17. Dependent claims 17, 36, and 55

The Office Action refers to Col. 4, lines 11-16 of Palmer as teaching the elements of the Appellants' claims 17, 36, and 55. In fact, the above cited section of Palmer merely states, “[t]he sum of the distances for a document is a confidence value or score. In a directory engine, the confidence value or score characterizes the confidence that a particular document falls within a particular category. In a search engine, the confidence value or score characterizes the relevance of a particular document to a given query.” However, there is nothing in the above language of

Palmer that teaches or reasonably suggests that each classifier or filter layer generates the relevancy score, which the Appellants' provide. It appears that the Examiner is not reading the claims in their totality (i.e., as a whole), but rather is arbitrarily selecting portions of the claims and attempting to create a link to portions of the prior art in a piecemeal manner. This conflicts with MPEP §§2116.01; 2141; 2141.02; and 2142 and 35 USC§103(a), and as such, the rejection is improper. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 17, 36, and 55.

18. Dependent claims 18, 37, and 56

The Office Action refers to Col. 4, lines 11-16 of Palmer as teaching the elements of the Appellants' claims 18, 37, and 56. In fact, the above cited section of Palmer merely states, “[t]he sum of the distances for a document is a confidence value or score. In a directory engine, the confidence value or score characterizes the confidence that a particular document falls within a particular category. In a search engine, the confidence value or score characterizes the relevance of a particular document to a given query.” However, there is nothing in the above language of Palmer that teaches or reasonably suggests that each classifier or filter layer generates the relevancy score, which the Appellants' provide. It appears that the Examiner is not reading the claims in their totality (i.e., as a whole), but rather is arbitrarily selecting portions of the claims and attempting to create a link to portions of the prior art in a piecemeal manner. This conflicts with MPEP §§2116.01; 2141; 2141.02; and 2142 and 35 USC§103(a), and as such, the rejection is improper. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to

reconsider and withdraw the rejections to dependent claims 18, 37, and 56.

19. Dependent claims 19, 38, and 57

The Office Action indicates that Col. 14, lines 29-35 of Palmer teaches the Appellants' "features of said feature vectors comprise words within a range of words located proximate to entities of interest in said document." The relevant portion of Palmer teaches, "[a] subset of the features that most strongly discriminate documents in one category from documents in another category are selected. For example, based on two million features, about one-half million most discriminative features or word phrases may be chosen. As shown in block 316, word vectors or feature vectors are constructed for each document in the expanded set. Each component of the feature vectors is the normalized value of the occurrence frequency of a particular feature in this document." However, nothing in the above quoted language in Palmer refers to the one-half million most discriminative feature or word phrases being located proximate to entities of interest in the document, which the Appellants' claims provide. As such, Palmer fails to teach all of the elements of the Appellants' claims 19, 38, and 57 rendering the rejection deficient. Furthermore, even a combination of Palmer with Woo fails to teach the entirety of the Appellants' claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw the rejections to dependent claims 19, 38, and 57.

E. CONCLUSION

In view of the foregoing, the Appellants respectfully submit that the collective cited prior art references, Palmer and Woo, do not teach or suggest the features defined by independent claims 1, 20, and 39, and as such, claims 1, 20, and 39 are patentable over Palmer in combination

with Woo. Further, dependent claims 2-19, 21-38, and 40-57 are similarly patentable over Palmer in combination with Woo, not only by virtue of their dependency from patentable independent claims, respectively, but also by virtue of the additional features of the Appellants' claimed invention they define. Thus, the Appellants respectfully request that the Board reconsider and withdraw the rejections of claims 1-57 and pass these claims to issue.

Please charge any deficiencies and credit any overpayments to Attorney's Deposit Account Number 09-0441.

Respectfully submitted,

Date: March 12, 2007

/Mohammad S. Rahman/

Mohammad S. Rahman, Esq.
Registration No. 43,029

Gibb & Rahman, LLC
2568-A Riva Road, Suite 304
Annapolis, MD, 21401
Voice: (301) 261-8625
Fax: (301) 261-8825
Customer No. 29154

IX. CLAIMS APPENDIX

1. A system for extracting information comprising:
 - a query input;
 - a database of documents;
 - a plurality of classifiers arranged in a hierarchical cascade of classifier layers, wherein each classifier comprises a set of weighted training data points comprising feature vectors representing any portion of a document, wherein each said feature vector is arranged only as a vector of counts for all features in a data point, and wherein said classifiers are operable to retrieve documents from said database based solely on whether said documents are relevant to said query input; and
 - a terminal classifier weighing an output from said cascade according to a rate of success of query terms being matched by each layer of said cascade.
2. The system of claim 1, wherein each classifier accepts an input distribution of data points and transforms said input distribution to an output distribution of said data points.
3. The system of claim 2, wherein each classifier is trained by weighing training data points at each classifier layer in said cascade by an output distribution generated by each previous classifier layer, and wherein weights of said training data points of said first classifier layer are uniform.
4. The system of claim 3, wherein each classifier is trained according to said query input.

5. The system of claim 2, wherein said query input is based on a minimum number of example documents.
6. The system of claim 1, wherein said document comprises data points comprising feature vectors representing any portion of said document.
7. The system of claim 1, wherein said documents comprise a file format capable of being represented by said feature vectors.
8. The system of claim 1, wherein said documents comprise any of text files, images, web pages, video files, and audio files.
9. The system of claim 2, wherein a classifier at each layer in said hierarchical cascade is trained for each layer with an expectation maximization methodology that maximizes a likelihood of a joint distribution of said training data points and latent variables.
10. The system of claim 9, wherein each layer of said cascade of classifiers is trained in succession from a previous layer by said expectation maximization methodology, wherein said output distribution is used as an input distribution for a succeeding layer.
11. The system of claim 9, wherein each layer of said cascade of classifiers is trained by successive iterations of said expectation maximization methodology until a convergence of

parameter values associated with said output distribution of each layer occurs in succession.

12. The system of claim 11, wherein said successive iterations comprise a fixed number of iterations.

13. The system of claim 9, wherein all layers of said cascade of classifiers are trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with output distributions of all layers occurs, wherein during each step of the of said iterations, the output distribution of each layer is used to weigh the input distribution of a succeeding layer.

14. The system of claim 13, wherein said successive iterations comprise a fixed number of iterations.

15. The system of claim 2, wherein each classifier layer generates a relevancy score associated with each a data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents.

16. The system of claim 2, wherein each classifier layer generates a relevancy score associated with said document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

17. The system of claim 5, wherein said terminal classifier generates a relevancy score

associated with each data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents, and wherein said relevancy score is computed by combining relevancy scores generated by classifiers at each layer of the cascade.

18. The system of claim 2, wherein said terminal classifier generates a relevancy score associated with a document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

19. The system of claim 1, wherein features of said feature vectors comprise words within a range of words located proximate to entities of interest in said document.

20. A method of extracting information, said method comprising:

- inputting a query;
- searching a database of documents based on said query;
- retrieving documents from said database based solely on whether said documents are relevant to said query using a plurality of classifiers arranged in a hierarchical cascade of classifier layers, wherein each classifier comprises a set of weighted training data points comprising feature vectors representing any portion of a document, wherein each said feature vector is arranged only as a vector of counts for all features in a data point; and
- weighing an output from said cascade according to a rate of success of query terms being matched by each layer of said cascade, wherein said weighing is performed using a terminal classifier.

21. The method of claim 20, wherein each classifier accepts an input distribution of data points and transforms said input distribution to an output distribution of said data points.
22. The method of claim 21, wherein each classifier is trained by weighing training data points at each classifier layer in said cascade by an output distribution generated by each previous classifier layer, and wherein weights of said training data points of said first classifier layer are uniform.
23. The method of claim 22, wherein each classifier is trained according to said query input.
24. The method of claim 21, wherein said query input is based on a minimum number of example documents.
25. The method of claim 20, wherein said document comprises data points comprising feature vectors representing any portion of said document.
26. The method of claim 20, wherein said documents comprise a file format capable of being represented by said feature vectors.
27. The method of claim 20, wherein said documents comprise any of text files, images, web pages, video files, and audio files.
28. The method of claim 21, wherein a classifier at each layer in said hierarchical cascade is

trained for each layer with an expectation maximization methodology that maximizes a likelihood of a joint distribution of said training data points and latent variables.

29. The method of claim 28, wherein each layer of said cascade of classifiers is trained in succession from a previous layer by said expectation maximization methodology, wherein said output distribution is used as an input distribution for a succeeding layer.

30. The method of claim 28, wherein each layer of said cascade of classifiers is trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with said output distribution of each layer occurs in succession.

31. The method of claim 30, wherein said successive iterations comprise a fixed number of iterations.

32. The method of claim 28, wherein all layers of said cascade of classifiers are trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with output distributions of all layers occurs, wherein during each step of the of said iterations, the output distribution of each layer is used to weigh the input distribution of a succeeding layer.

33. The method of claim 32, wherein said successive iterations comprise a fixed number of iterations.

34. The method of claim 21, wherein each classifier layer generates a relevancy score associated with each a data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents.

35. The method of claim 21, wherein each classifier layer generates a relevancy score associated with said document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

36. The method of claim 24, wherein said terminal classifier generates a relevancy score associated with each data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents, and wherein said relevancy score is computed by combining relevancy scores generated by classifiers at each layer of the cascade.

37. The method of claim 21, wherein said terminal classifier generates a relevancy score associated with a document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

38. The method of claim 20, wherein features of said feature vectors comprise words within a range of words located proximate to entities of interest in said document.

39. A program storage device readable by computer, tangibly embodying a program of instructions executable by said computer to perform a method of extracting information, said method comprising:

inputting a query;

searching a database of documents based on said query;

retrieving documents from said database based solely on whether said documents are relevant to said query using a plurality of classifiers arranged in a hierarchical cascade of classifier layers, wherein each classifier comprises a set of weighted training data points comprising feature vectors representing any portion of a document, wherein each said feature vector is arranged only as a vector of counts for all features in a data point; and

weighing an output from said cascade according to a rate of success of query terms being matched by each layer of said cascade, wherein said weighing is performed using a terminal classifier.

40. The program storage device of claim 19, wherein each classifier accepts an input distribution of data points and transforms said input distribution to an output distribution of said data points.

41. The program storage device of claim 40, wherein each classifier is trained by weighing training data points at each classifier layer in said cascade by an output distribution generated by each previous classifier layer, and wherein weights of said training data points of said first classifier layer are uniform.

42. The program storage device of claim 41, wherein each classifier is trained according to said query input.

43. The program storage device of claim 40, wherein said query input is based on a minimum number of example documents.

44. The program storage device of claim 39, wherein said document comprises data points comprising feature vectors representing any portion of said document.

45. The program storage device of claim 39, wherein said documents comprise a file format capable of being represented by said feature vectors.

46. The program storage device of claim 39, wherein said documents comprise any of text files, images, web pages, video files, and audio files.

47. The program storage device of claim 40, wherein a classifier at each layer in said hierarchical cascade is trained for each layer with an expectation maximization methodology that maximizes a likelihood of a joint distribution of said training data points and latent variables.

48. The program storage device of claim 47, wherein each layer of said cascade of classifiers is trained in succession from a previous layer by said expectation maximization methodology, wherein said output distribution is used as an input distribution for a succeeding layer.

49. The program storage device of claim 47, wherein each layer of said cascade of classifiers is trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with said output distribution of each layer occurs in

succession.

50. The program storage device of claim 49, wherein said successive iterations comprise a fixed number of iterations.

51. The program storage device of claim 47, wherein all layers of said cascade of classifiers are trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with output distributions of all layers occurs, wherein during each step of the of said iterations, the output distribution of each layer is used to weigh the input distribution of a succeeding layer.

52. The program storage device of claim 51, wherein said successive iterations comprise a fixed number of iterations.

53. The program storage device of claim 40, wherein each classifier layer generates a relevancy score associated with each a data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents.

54. The program storage device of claim 40, wherein each classifier layer generates a relevancy score associated with said document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

55. The program storage device of claim 43, wherein said terminal classifier generates a

relevancy score associated with each data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents, and wherein said relevancy score is computed by combining relevancy scores generated by classifiers at each layer of the cascade.

56. The program storage device of claim 40, wherein said terminal classifier generates a relevancy score associated with a document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

57. The program storage device of claim 39, wherein features of said feature vectors comprise words within a range of words located proximate to entities of interest in said document.

X. EVIDENCE APPENDIX

There is no other evidence known to Appellants, Appellants' legal representative or Assignee which would directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

XI. RELATED PROCEEDINGS APPENDIX

There are no other related proceedings known to Appellants, Appellants' legal representative or Assignee which would directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.